

4-18-2007

## Semantic Web Applications in Industry, Government, Health Care and Life Sciences

Amit P. Sheth

*Wright State University - Main Campus, amit@sc.edu*

Follow this and additional works at: <https://corescholar.libraries.wright.edu/knoesis>



Part of the [Bioinformatics Commons](#), [Communication Technology and New Media Commons](#), [Databases and Information Systems Commons](#), [OS and Networks Commons](#), and the [Science and Technology Studies Commons](#)

---

### Repository Citation

Sheth, A. P. (2007). Semantic Web Applications in Industry, Government, Health Care and Life Sciences. .  
<https://corescholar.libraries.wright.edu/knoesis/31>

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).



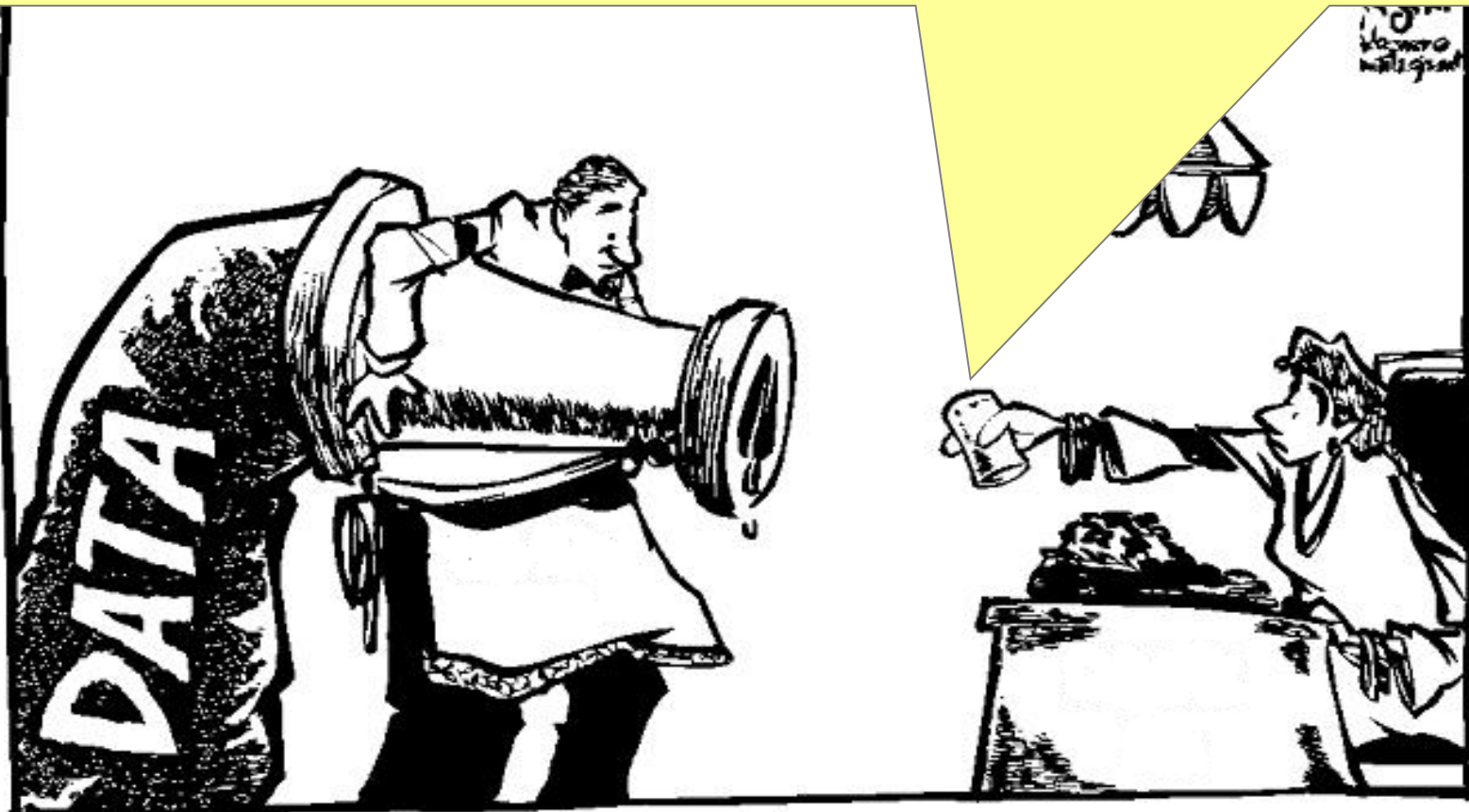
COLLECTING THE DOTS | CONNECTING THE DOTS

# Semantic Web applications in Industry, Government, Health care and Life Sciences

Amit Sheth

- Death by data: amount, variety, sources
- How to exploit this data?
- Silver Bullet – SEMANTICS (approach) & Semantic Web (technology)
- Applications to show the value
  - Industry (financial services)
  - Government (intelligence)
  - Science (health care and biomedicine)

Not data (search), but **integration**, analysis and **insight**, leading to **decisions** and **discovery**



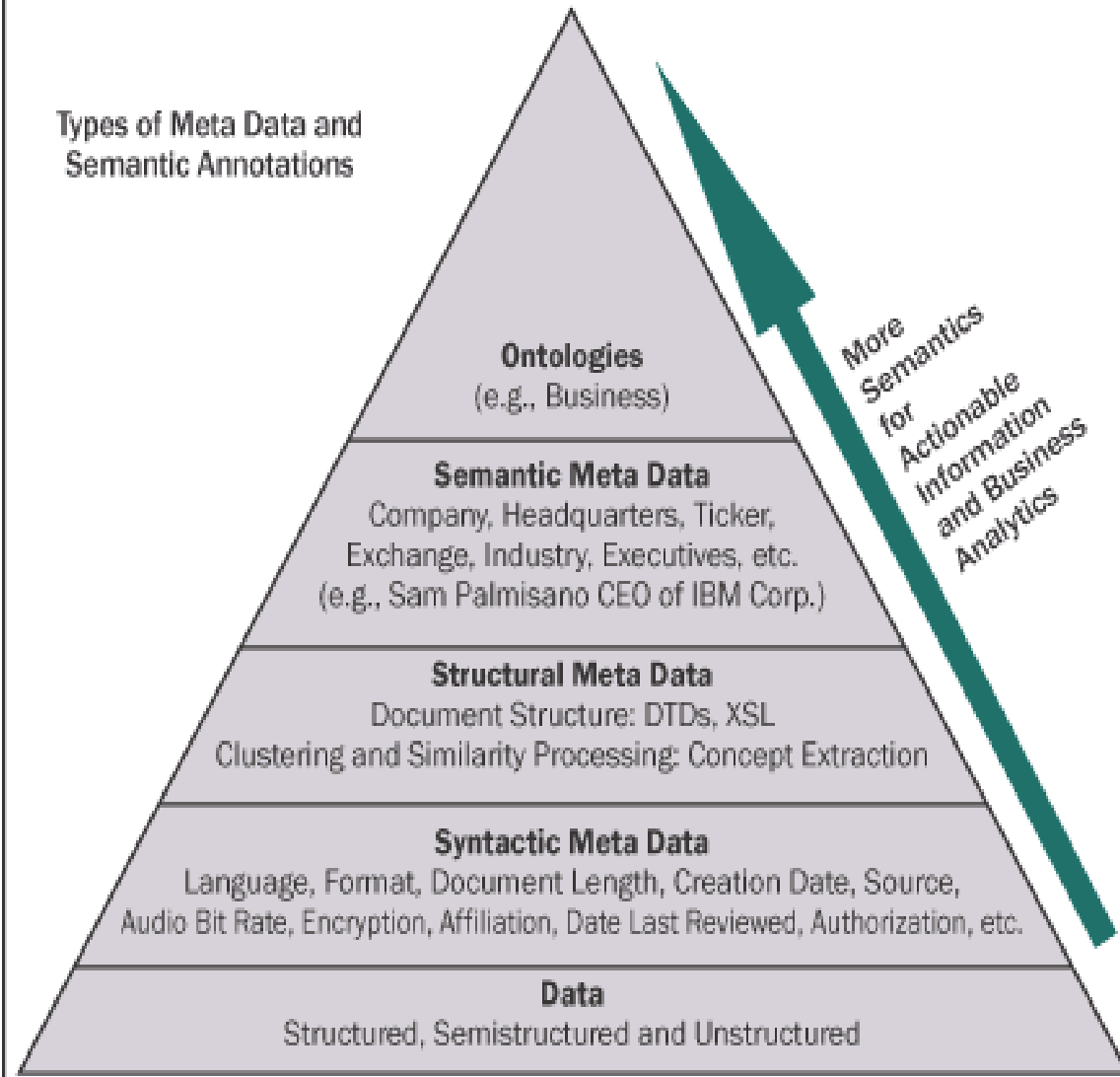
- Data captured per year = 1 exabyte ( $10^{18}$ )  
(Eric Neumann, Science, 2005)
- How much is that?
  - Compare it to the estimate of the total words ever spoken by humans = 12 exabyte
- Continued high cost of interoperability and integration
- Needle in the haystack

- Multiple formats: Structured, unstructured, semi-structured
- Multimodal: text, image, a/v, sensor, scientific/engineering
- Thematic, Spatial, Temporal
- Enterprise to Globally Distributed

*What?*

## Moving from Syntax/Structure to Semantics

Types of Meta Data and  
Semantic Annotations





Semantics: **Meaning & Use of Data**

Semantic Web: **Labeling data on the Web so both humans and machines can use them more effectively**

i.e., Formal, machine processable  
description  $\Rightarrow$  more automation;  
emerging standards/technologies

(RDF, OWL, Rules, ...)

# Semantic Annotation - document

Blue-chip bonanza continues

company company company  
Dow above 9,000 as [HP](#), [Home Depot](#) lead advance; [Microsoft](#) upgrade helps techs.

date time  
August 22, 2002: 11:44 AM EDT

phrase phrase  
By Alexandra Twin, CNN/Money Staff Writer

city company  
[New York](#) (CNN/Money) - An upgrade of software leader [Microsoft](#) and strength in blue chips including

company company weekday  
[Hewlett-Packard](#) and [Home Depot](#) were among the factors pushing stocks higher at midday Thursday,

financial index  
with the [Dow Jones industrial average](#) spending time above the 9,000 level.

time financial index  
Around 11:40 a.m. ET, the [Dow Jones industrial average](#) gained 65.06 to 9,022.09, continuing a more

date stock exchange  
than 1,300-point resurgence since July 23. The [Nasdaq](#) composite gained 9.12 to 1,418.37.

financial index  
[The Standard & Poor's 500 index](#) rose 9.61 to 958.97.

company stockSym \$ \$  
[Hewlett-Packard](#) ( [HPQ](#): up \$0.33 to \$15.03, Research, Estimates) said a report shows its share of

continent region continent  
the printer market grew in the second quarter, although another report showed that its share of the

continent  
computer server market declined in [Europe](#), the [Middle East](#) and [Africa](#).

company stockSym \$ \$  
[Home Depot](#) ( [HD](#): up \$1.07 to \$33.75, Research, Estimates) was up for the third straight day after

topping fiscal second-quarter earnings estimates on Tuesday.

tech category company  
Tech stocks managed a turnaround. [Software](#) continued to rise after [Salomon Smith Barney](#) upgraded

company stockSym \$ \$  
No. 1 software maker [Microsoft](#) ( [MSFT](#): up \$0.55 to \$52.83, Research, Estimates) to "outperform"

company  
from "neutral" and raised its price target to \$59 from \$56. Business software makers [Oracle](#)

stockSym \$ \$ company stockSym \$ \$  
( [ORCL](#): up \$0.18 to \$10.94, Research, Estimates), [PeopleSoft](#) ( [PSFT](#): up \$1.17 to \$20.67,

company stockSym \$ \$  
Research, Estimates) and [BEA Systems](#) ( [BEAS](#): up \$0.28 to \$7.12, Research, Estimates)

all rose in tandem.

competes with

*How?*

**Ontology:** Agreement with Common Vocabulary & Domain Knowledge

**Semantic Annotation:** metadata (manual & automatic metadata extraction)

**Reasoning:** semantics enabled search, integration, analysis, mining, discovery

- Agreement
- Common Nomenclature
- Conceptual Model with associated knowledgebase (ground truth/facts) for an industry, market, field of science, activity
- In some domains, extensive building of open-source ontologies, in others, build as you go

# Semantic Annotation – news feed

`<Entity id="494805" class="DrugOntology#Organization"`  
Public Health Advisory **FDA** Announces Important Change  
`<Entity id="492805" class="DrugOntology#prescription_`  
Selective and Non-Selective Non-Steroidal Anti-Inflamma  
`class="DrugOntology#prescription_drug_generic">NSAID`

Today, the Food and Drug Administration (FDA) is announce  
`<Entity id="474305" class="DrugOntology#Organization"`  
to voluntarily withdraw `<Entity id="122805" class="`  
`DrugOntology#prescription_drug_brandname">Bextra`  
`</Entity>` (`<Entity id="10288" class="DrugOntology#pre`  
`valdecoxib</Entity>`) from the market. **Pfizer** has agreed to  
of **Bextra** in the `<Entity id="7852" class="Sweto#country"`  
further discussions with the agency.

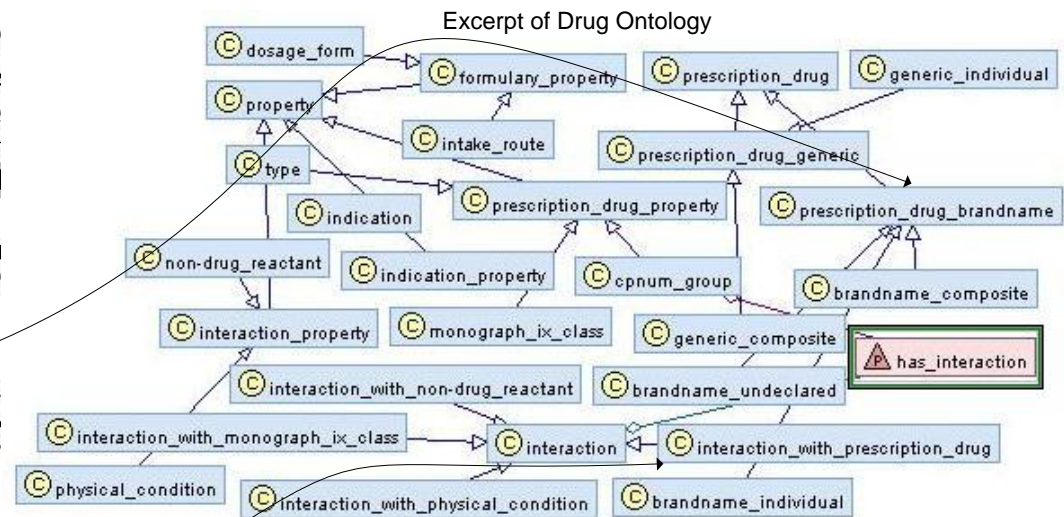
.....

This request is based on:

-Reports of serious and potentially life-threatening skin reactions, including deaths,  
in patients using **Bextra**. The risk of these reactions in individual patients is unpredictable,  
occurring in patients with and without a prior history of `<Entity id="14280" class="`  
`DrugOntology#interaction_with_physical_condition>sulfa allergy`  
`</Entity>`

.....

Date created: `<Regex type="date">April 7, 2005</Regex>`



## Sample Created Metadata

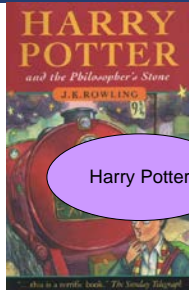
```
<Entity id="122805"
class="DrugOntology#prescription_drug_brandname">
Bextra
<Relationship id="442134"
class="DrugOntology#has_interaction">
<Entity id="14280" class="DrugOntology
#interaction_with_physical_condition">sulfa allergy
</Entity>
</Relationship>
</Entity>
```

Joe, a high school student researching the pseudo-history behind *The Da Vinci Code*. Teacher suggests him to focus on:

- The Last Supper by **Leonardo da Vinci**
- Et in Arcadia ego by **Nicolas Poussin**

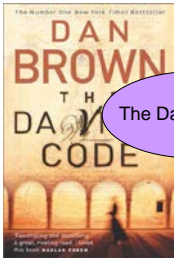
It so happens that Joe's younger brother is reading *Harry Potter and the Philosopher's Stone* by J.K. Rowling. A casual conversation between the brothers leads to a discussion about **Nicolas Flamel**, who in the Harry Potter novel is said to have possessed the philosopher's stone that could convert base metals to gold. At this point their father, an amateur historian suggests that there might have been a real person by the name Nicholas Flamel, and the three wonder if there is any **link** between these seemingly unrelated stories.

# Anecdotal Example



Harry Potter

UNDISCOVERED PUBLIC KNOWLEDGE  
Discovering connections hidden in text



The Da Vinci code



Et in Arcadia Ego



Santa Maria delle Grazie



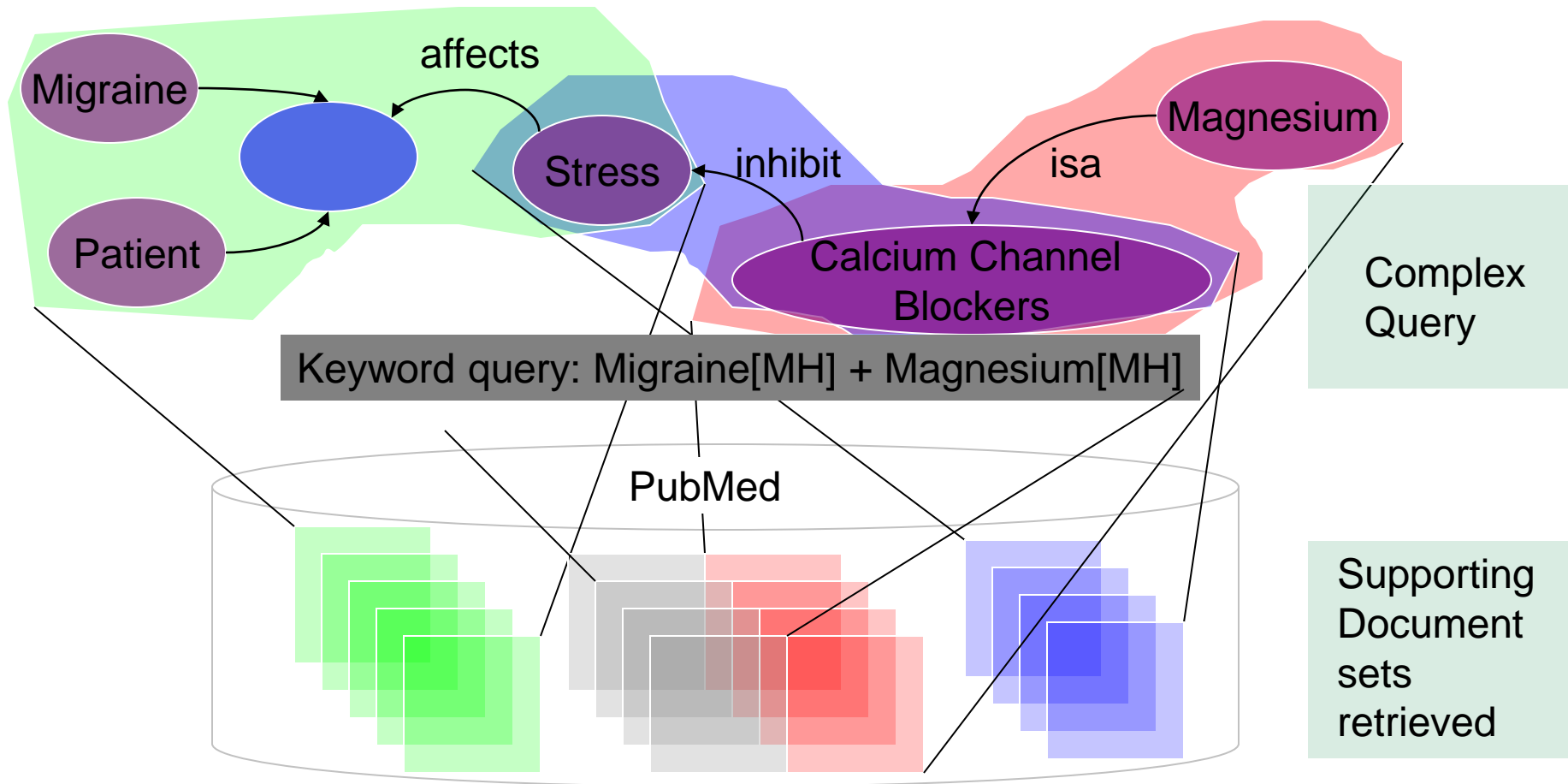
The Priority of Sion was supposedly led by a [Grand Master](#) or [Nautonnier](#).

1. [Ugo de Blancheford](#) (1150-1151)
2. [Bernard de Tremblay](#) (1151-1153)
3. [Guillaume de Chanaleilles](#) (1153-1154)
4. [Evrard de N...?](#) (1154-1154)
5. [Andr  de Montbard](#) (1155-1156)
6. [Bertrand de Blanchefort](#) (1156-1169)
7. [Philippe de Milly](#) (1169-1170)
8. [Eudes de Saint-Amand](#) (1170-1180)
9. [Arnaud de Toroge](#) (1181-1184)
10. [G rard de Rideford](#) (1184-1188)
11. [Jean de Gisors](#) (1188-1220)
12. [Marie de Saint-Clair](#) (1220-1266)
13. [Guillaume de Gisors](#) (1266-1307)
14. [Edouard de Bar](#) (1307-1336)
15. [Jeanne de Bar](#) (1336-1351)
16. [Jean de Saint-Clair](#) (1351-1366)
17. [Blanche d'Evreux](#) (1366-1398)
18. [Nicolas Flamel](#) (1398-1418)
19. [Rene d'Anjou](#) (1418-1480)
20. [Iolande de Bar](#) (1480-1483)
21. [Sandro Filipepi AKA Botticelli](#) (1483-1510)
22. [Leonardo da Vinci](#) (1510-1519)
23. [Charles III \(Duke of Bourbon-Montpensier\)](#) (1519-1527)
24. [Ferdinand de Gonzague](#) (1527-1556)
25. [Michel de Notre-Dame AKA Nostradamus](#) (1556-1566)
26. [Duc de Longueville & Nicolas Froumenteau](#) (1566-1575)
27. [Louis de Nevers](#) (1575-1595)
28. [Robert Fludd](#) (1595-1637)
29. [Johann Valentin Andrea](#) (1637-1654)
30. [Robert Boyle](#) (1654-1691)
31. [Isaac Newton](#) (1691-1727)
32. [Charles Radclyffe](#) (1727-1746)
33. [Charles de Lorraine](#) (1746-1780)
34. [Maximillian de Lorraine](#) (1780-1801)
35. [Charles Nodier](#) (1801-1844)
36. [Victor Hugo](#) (1844-1885)
37. [Claude Debussy](#) (1885-1918)
38. [Jean Cocteau](#) (1918-1963)
39. [Pierre Plantard](#) (1963-1981)

A second List of the Grand Masters of the Priority of Sion that included the names of Roger Patrice Pelat and Thomas Plantard appeared in 1989, but it should not be confused with the



# Hypothesis Driven retrieval of Scientific Literature



**Aim:** Legislation (PATRIOT ACT) requires banks to identify ‘who’ they are doing business with

## Problem

- Volume of internal and external data needed to be accessed
- Complex name matching and disambiguation criteria
- Requirement to ‘risk score’ certain attributes of this data

## Approach

- Creation of a ‘risk ontology’ populated from trusted sources (OFAC etc); Sophisticated entity disambiguation
- Semantic querying, Rules specification & processing

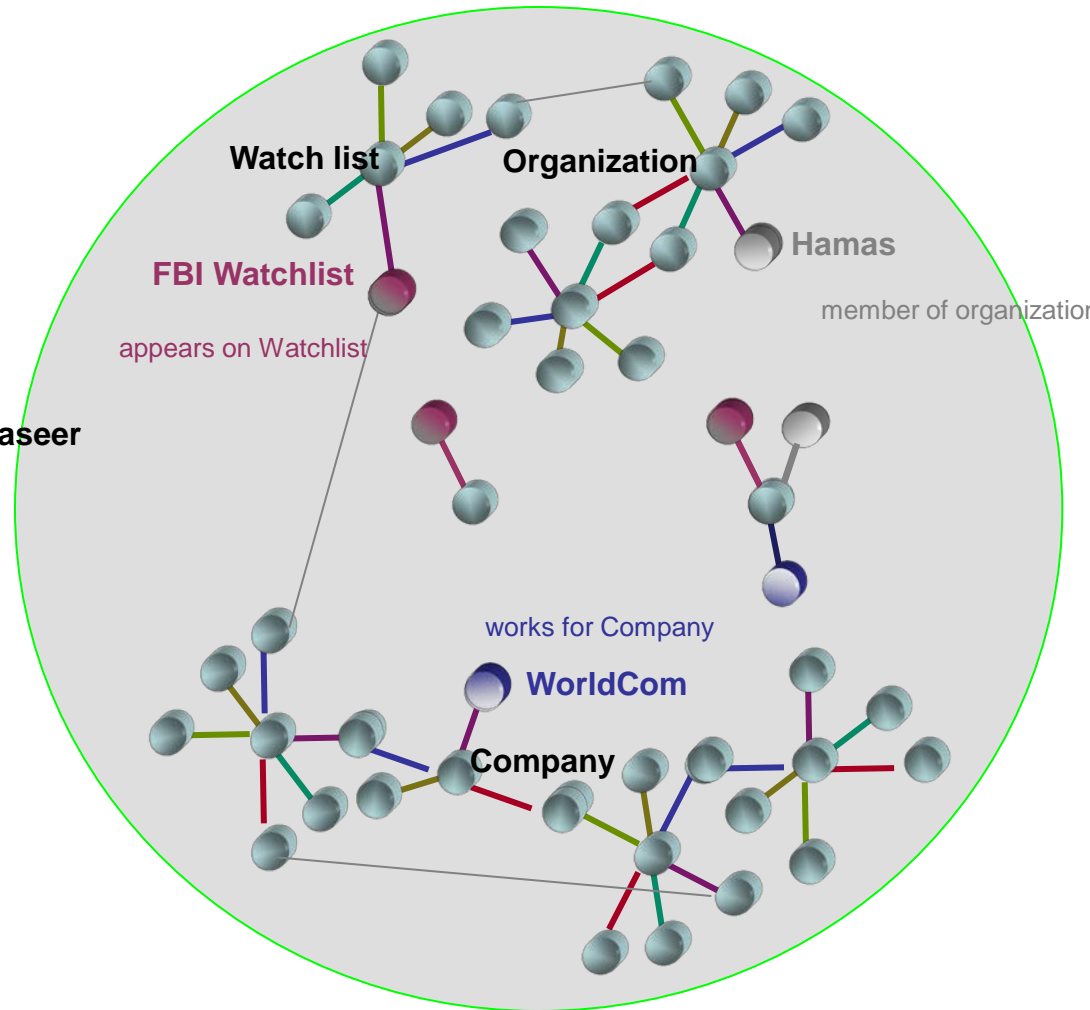
## Solution

- Rapid and accurate KYC checks
- Risk scoring of relationships allowing for prioritisation of results; Full visibility of sources and trustworthiness

# The Process

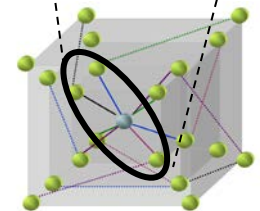
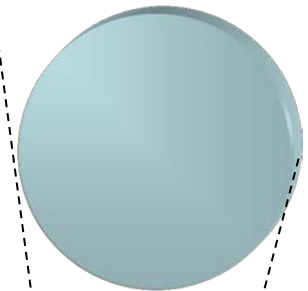


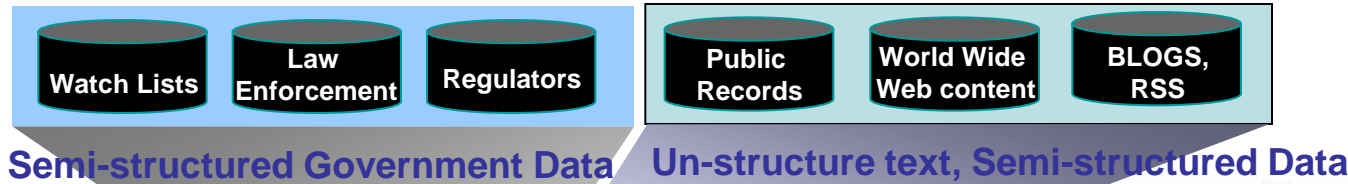
Ahmed Yaseer



## Ahmed Yaseer:

- **Appears on Watchlist 'FBI'**
- **Works for Company 'WorldCom'**
- **Member of organization 'Hamas'**





Establishing  
New Account

**PIRAS**  
Customer Identification and Risk Assessment

**Check Individual** | Check Organization | Status & Results | Reporting

Name: [First Name] [Last Name] | Primary Residence Address: [Street] [City] [State/Province] [Zip/Postal Code] [Country]

Date of Birth: [MM/DD/YYYY] | Submitter: [Name] [Address] [City] [State/Province] [Zip/Postal Code] [Country]

Individual Attributes: [All Fields] [Selected Fields] [None]

Banking Center: [Country] [City] [State/Province] [Zip/Postal Code]



Example of  
Fraud Prevention  
application used in  
financial services

**Individual Check - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

**PIRAS**  
Customer Identification and Risk Assessment

User: Larry Parker | [Sign Off](#)

[Check Individual](#) | [Check Organization](#) | [Status & Results](#) | [Reporting](#)

**LMZ Shipping, LLC** ENTITY  
Check ID: 14092 Status: Pending — [Cancel or Refer this Record](#)

Score: 65 | [Print \(PDF\)](#)  
Enter explanation of score...

[Summary](#) | [Match Details](#) | [WebFountain](#) | [Media Content](#) | [Attributes](#) | [Associations](#)

[Launch Associations Visualizer](#)

<b>Is Related to</b>	<a href="#">Wojtech Moroski</a> Wojtech Moroski is related to Rabbita Trust Rabbita Trust appears on FBI Watch List
<b>Undertakes</b>	12 Thompson Av., Marietta, Georgia
<b>Is related to</b>	Wojtech Moroski
<b>Active in</b>	Washington
<b>Active in</b>	US

[Perform a New Check](#)

**SEMAG!X**  
POWER • THROUGH • RELEVANCE

[Terms, conditions, caveats and small print](#)

© Copyright 2004. Semagix. All Rights Reserved.

## Active Semantic Medical Records

(operational since January 2006)

### Goals:

- Increase efficiency
- Reduce Errors, Improve Patient Satisfaction & Reporting
- Improve Profitability (better billing)

### Technologies:

- Ontologies, semantic annotations & rules
- Service Oriented Architecture

- Semantic Browser: contextual browsing of PubMed

- Passenger Threat Analysis
- Need to Know -> Demo
- Financial Irregularity \*

\* a classified application

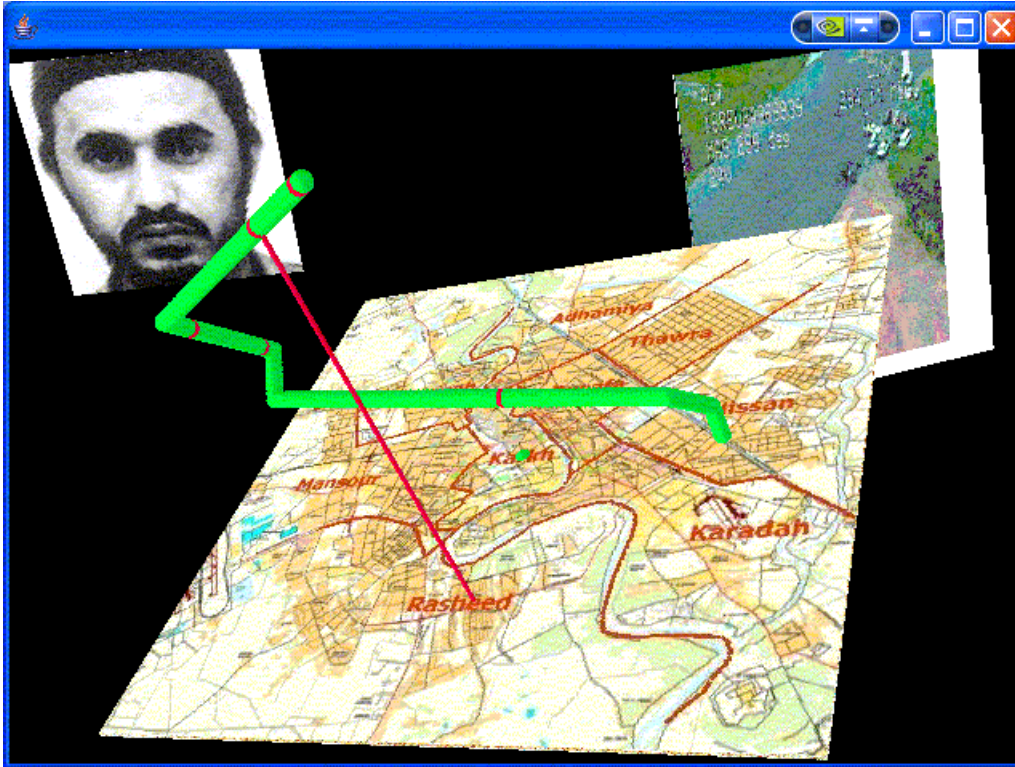
## Aim

- Visualization with interactive search and analytics interface

## Problem (examples)

- Need for intuitive visual display of **semantic analytics** showing "connections between the dots" between heterogeneous documents and multi-modal content
- Need for graphical tracking and association of activities to discover semantic associations between events using thematic and topological relations





- Visualization of association unfolding over time
- Integration of associated multimedia content
- Separate **Temporal**, **Geospatial**, and **Thematic** ontologies describe data
- [DEMO](#)

- World class research center- coupled with daytaOhio for tech transfer and commercialization
- Core expertise in
  - data management: integration, mining, analytics, visualization
  - distributed computing: services/grid computing
  - Semantic Web
  - Bioinformatics, etc.
- With domain/application expertise in Government, Industry, Biomedicine
- Member of World Wide Web Consortium and extensive industry relationships

- Interns, future employees
- Targets research & prototyping
- dataOhio supported technology transfer/incubator
- Joint projects (e.g., C3 funding from DoD)
- Guidance on using standards and technologies

“increasing the value of data to you and your customers”

# Introducing



<http://knoesis.wright.edu>